UNITED STATES PATENT AND TRADEMARK OFFICE

UNITED STATES DEPARTMENT OF COMMERCE
United States Patent and Trademark Office
Address: COMMISSIONER FOR PATENTS
P.O. Box 1450
Alexandria, Virginia 22313-1450
www.uspto.gov

APPLICATION NO.	FILING DATE	FIRST NAMED INVENTOR	ATTORNEY DOCKET NO.	CONFIRMATION NO.
09/826,715	04/05/2001	Chang-Qing Shu	00-4023	3951

32127 7590 04/22/2004

VERIZON CORPORATE SERVICES GROUP INC.
C/O CHRISTIAN R. ANDERSEN
600 HIDDEN RIDGE DRIVE
MAILCODE HQEO3H14
IRVING, TX 75038

EXAMINER	
BRANT, DMITRY	
ART UNIT	**PAPER NUMBER**
2655	4

DATE MAILED: 04/22/2004

Please find below and/or attached an Office communication concerning this application or proceeding.

Office Action Summary	Application No.	Applicant(s)
	09/826,715	SHU ET AL.
	Examiner	Art Unit
	Dmitry Brant	2655

-- The MAILING DATE of this communication appears on the cover sheet with the correspondence address --

Period for Reply

A SHORTENED STATUTORY PERIOD FOR REPLY IS SET TO EXPIRE 3 MONTH(S) FROM THE MAILING DATE OF THIS COMMUNICATION.

- Extensions of time may be available under the provisions of 37 CFR 1.136(a). In no event, however, may a reply be timely filed after SIX (6) MONTHS from the mailing date of this communication.
- If the period for reply specified above is less than thirty (30) days, a reply within the statutory minimum of thirty (30) days will be considered timely.
- If NO period for reply is specified above, the maximum statutory period will apply and will expire SIX (6) MONTHS from the mailing date of this communication.
- Failure to reply within the set or extended period for reply will, by statute, cause the application to become ABANDONED (35 U.S.C. § 133). Any reply received by the Office later than three months after the mailing date of this communication, even if timely filed, may reduce any earned patent term adjustment. See 37 CFR 1.704(b).

Status

1)☒ Responsive to communication(s) filed on 4/05/01.

2a)☐ This action is **FINAL**. 2b)☒ This action is non-final.

3)☐ Since this application is in condition for allowance except for formal matters, prosecution as to the merits is closed in accordance with the practice under *Ex parte Quayle*, 1935 C.D. 11, 453 O.G. 213.

Disposition of Claims

4)☒ Claim(s) 1-32 is/are pending in the application.

4a) Of the above claim(s) _____ is/are withdrawn from consideration.

5)☐ Claim(s) _____ is/are allowed.

6)☒ Claim(s) 1-32 is/are rejected.

7)☐ Claim(s) _____ is/are objected to.

8)☐ Claim(s) _____ are subject to restriction and/or election requirement.

Application Papers

9)☐ The specification is objected to by the Examiner.

10)☐ The drawing(s) filed on _____ is/are: a)☐ accepted or b)☐ objected to by the Examiner.

Applicant may not request that any objection to the drawing(s) be held in abeyance. See 37 CFR 1.85(a).

Replacement drawing sheet(s) including the correction is required if the drawing(s) is objected to. See 37 CFR 1.121(d).

11)☐ The oath or declaration is objected to by the Examiner. Note the attached Office Action or form PTO-152.

Priority under 35 U.S.C. § 119

12)☐ Acknowledgment is made of a claim for foreign priority under 35 U.S.C. § 119(a)-(d) or (f).

a)☐ All b)☐ Some * c)☐ None of:

1.☐ Certified copies of the priority documents have been received.

2.☐ Certified copies of the priority documents have been received in Application No. _____.

3.☐ Copies of the certified copies of the priority documents have been received in this National Stage application from the International Bureau (PCT Rule 17.2(a)).

* See the attached detailed Office action for a list of the certified copies not received.

Attachment(s)

1)☒ Notice of References Cited (PTO-892)

2)☐ Notice of Draftsperson's Patent Drawing Review (PTO-948)

3)☒ Information Disclosure Statement(s) (PTO-1449 or PTO/SB/08) Paper No(s)/Mail Date 2.

4)☐ Interview Summary (PTO-413) Paper No(s)/Mail Date. _____.

5)☐ Notice of Informal Patent Application (PTO-152)

6)☐ Other: _____.

DETAILED ACTION

Claim Rejections - 35 USC § 102

1. The following is a quotation of the appropriate paragraphs of 35 U.S.C. 102 that form the basis for the rejections under this section made in this Office action:

> A person shall be entitled to a patent unless –
>
> (b) the invention was patented or described in a printed publication in this or a foreign country or in public use or on sale in this country, more than one year prior to the date of application for patent in the United States.

2. Claims 1 and 6 are rejected under 35 U.S.C. 102(b) as being anticipated by Sharma et al (5,862,519)

The U.S. patent of Sharma et al. teach computer-based apparatus (system) and hence the methods and computer code necessary to implement this system are inherent ~~inevitably~~ part of their teachings.

Limitations	Sharma et al.
• Receiving frames of acoustic data • Determining cepstral coefficients for each of the received frames of acoustic data	Computing Spectral Variation Function (SVF), which is based on the Euclidian norm of delta cepstral coefficients of each individual frame (**Col. 6, lines 32-38**). This process inherently presupposes *fragmentation* of acoustic data into frames and computing cepstral coefficients for each frame.
• Segmenting the received frames of acoustic data based on the determined cepstral coefficients.	Segmentation is performed based on the values of Kmax and Kmin (**elems. 20, 22, FIG. 1 and Col. 7, lines 23-28**). K max (maximum number of segments for each frame) is computed using SVF (**Col. 6, line 29-31**), which itself is a function of cepstral coefficients (**Eq. 2**). Therefore, segmentation of each frame is ultimately performed based on the cepstral coefficients.

3. Claims 10 and 12-14, 15, 18-19, 20, 22-25, 28, 29 are rejected under 35 U.S.C. 102(b) as being anticipated by Juang et al. (5,812,972)

The U.S. patent of Juang et al. teach computer-based apparatus (system) and hence the methods and computer code necessary to implement this system are inevitably part of their teachings.

Claim#	Limitations	Juang et al.
10,15	Receiving frames of acoustic data Determining segmentation information corresponding to the received frames of acoustic data Determining at least one weighting parameter based on the determined segmentation information Recognizing patterns in the received frames of acoustic data using the at least one weighting parameter	**(Elem. 315, FIG. 3)**, which corresponds to a frame of acoustic data. **(Elem. 330, FIG. 3)** **(Elem. 360, FIG. 3)**, where W(i) is a weighting factor based on the confidence level of the current segmentation vector. **(Col. 8, lines 55-59)** **(Elem. 355, FIG. 3)**
12,17	Determining , based on the frames of acoustic data, recognition hypothesis scores using a Hidden Markov Model	**(Elem. 327, FIG. 3)** uses HMM to determine the next likely state **(Col. 7, lines 44-48)**
13, 14, 19	Modifying the recognition hypothesis scores based on the at least one weighting parameter and the recognizing patterns in the frames of acoustic data further uses the modified recognition hypothesis scores	As it can be seen from **FIG. 3**, readjusted scores are fed from block **360** back into block **325**. Thus the HMM and corresponding scores are updated using the weighting factor Wi **(elem. 360, FIG. 3)** which reflects the level of confidence in the current model. Therefore, the final recognition result **(Elem. 355, FIG. 3)** will incorporate the modification through weighting factor W.

20, 29	Receiving frames of acoustic data Determining first segmentation information corresponding to the received frames of acoustic data and second segmentation information corresponding to the received frames of acoustic data. Determining at least one weighting parameter based on the determined second segmentation information Recognizing patterns in the received frames of acoustic data using the at least one weighting parameter.	**(Elem. 315, FIG. 3)**, which corresponds to a frame of acoustic data. For each sequential frame, modified Observation vector O''(i), is computed along with segmentation vector A(i) **(Col. 8, lines 7-17)** **(Elem. 360, FIG. 3)**, where W(i) is a weighting factor based on the confidence level of the current segmentation vector. **(Col. 8, lines 55-59)** **(Elem. 355, FIG. 3)**
22	Comparing the determined first and second segmentation information	Col. 8, lines 45 - 48

23	The recognizing patterns in the frames of acoustic data is based on the comparison of the first and second segmentation information	As it can be seen from **FIG. 3**, readjusted scores are fed from block **360** back into block **325**. Thus the HMM and corresponding scores are updated using the weighting factor Wi **(elem. 360, FIG. 3)** which reflects the level of confidence in the current model. Therefore, the final recognition result **(Elem. 355, FIG. 3)** will incorporate the modification through weighting factor W which includes comparison of O''(i) and A(i)
24	Determining , based on the frames of acoustic data, recognition hypothesis scores using a Hidden Markov Model	**(Elem. 327, FIG. 3)** uses HMM to determine the next likely state **(Col. 7, lines 44-48)**
25,28	Modifying the recognition hypothesis scores based on the at least one weighting parameter and the recognizing patterns in the frames of acoustic data further uses the modified recognition hypothesis scores	As it can be seen from **FIG. 3**, readjusted scores are fed from block **360** back into block **325**. Thus the HMM and corresponding scores are updated using the weighting factor Wi **(elem. 360, FIG. 3)** which reflects the level of confidence in the current model. Therefore, the final recognition result **(Elem. 355, FIG. 3)** will incorporate the modification through weighting factor W.

Claim Rejections - 35 USC § 103

3. The following is a quotation of 35 U.S.C. 103(a) which forms the basis for all obviousness rejections set forth in this Office action:

> (a) A patent may not be obtained though the invention is not identically disclosed or described as set forth in section 102 of this title, if the differences between the subject matter sought to be patented and the prior art are such that the subject matter as a whole would have been obvious at the time the

invention was made to a person having ordinary skill in the art to which said subject matter pertains. Patentability shall not be negatived by the manner in which the invention was made.

4. Claims 2-5, 7-9, 30, are rejected under 35 U.S.C. 103(a) as being unpatentable over Sharma et al.

The U.S. patent of Sharma et al. teach computer-based apparatus (system) and hence the methods and computer code, such as data structures, necessary to implement this system are inherent ~~inevitably~~ part of their teachings.

As per claims 2-4, Sharma et al. do not disclose determining the number of peaks in cepstral coefficients for each frame and then segmenting the frames based on the comparison of the number of peaks in each of the sequential frames **(Peak N Difference)**. The applicant discloses that the reason for tracking the variation in the number of peaks is to identify frames having phoneme boundaries, where the number of cepstral coefficient peaks change rapidly [Disclosure, Page. 6, lines 15-16].

Similarly, Sharma et al. applies this principle for computing the maximum number of peaks (Kmax). Sharma et al. discloses Spectral Variation (SVF) function which is computed based on the time variation of cepstral coefficients for each frame. (Col. 6, lines 32-38). To identify segments based on phoneme boundaries, Sharma tracks the peaks in SVF, because SVF exhibits peaks at boundaries where characteristics of speech change rapidly (Col. 6, lines 63-67). SVF tracks the frame-to-frame changes between the corresponding cepstral coefficients within individual frames (See Eq. 2) and thus, changes in the number of peaks would also affect SVF, so that SVF would

track the frames identified by the applicant's method as having high "**Peak_N_Difference**".

Therefore, it would have been obvious to one of ordinary skill in the art at the time the invention was made to modify Sharma to use the difference in the number of cepstral peaks instead of SVF, because these methods track the same cepstral properties in a similar fashion. Computing the difference in the number of peaks is a variation of SVF, and while not being as exact as SVF, it has the advantage of computational simplicity.

As per claim 5, Sharma et al. discloses <u>receiving frames of acoustic data and computing cepstral coefficients for each frame</u> - Sharma et al. computes Spectral Variation Function (SVF), which is based on the Euclidian norm of delta cepstral coefficients of each individual frame (**Col. 6, lines 32-38**). This process inherently presupposes fragmentation of acoustic data into frames and computing cepstral coefficients for each frame.

Sharma et al. do not disclose a processing unit that determines <u>the number of peaks</u> in cepstral coefficients for each frame and then <u>segments the frames</u> <u>based on the comparison of the number of peaks in each of the sequential frames</u>

However, Sharma et al.'s Spectral Variation (SVF) function is computed based on the time variation of cepstral coefficients for each frame. (Col. 6, lines 32-38). To identify segments based on phoneme boundaries, Sharma tracks the peaks in SVF, because SVF exhibits peaks at boundaries where characteristics of speech change

rapidly (Col. 6, lines 63-67). SVF tracks the frame-to-frame changes between the corresponding cepstral coefficients within individual frames (See Eq. 2) and thus, changes in the number of peaks would also affect SVF, so that SVF would track the frames identified by the applicant's method as having high "**Peak_N_Difference**".

Therefore, it would have been obvious to one of ordinary skill in the art at the time the invention was made to modify Sharma et al. to use the difference in the number of cepstral peaks instead of SVF, because these methods track the same cepstral properties in a similar fashion. Computing the difference in the number of peaks is a variation of SVF, and while not being as exact as SVF, it has the advantage of computational simplicity.

5. Claims 11,16, 18, 22, 26-27, are rejected under 35 U.S.C. 103(a) as being unpatentable over Juang et al., in view of Sharma et al.

As per claims 11 and 16, Juang et al. do not disclose "determining cepstral coefficients for the received frames of acoustic data, wherein the determining of the segmentation information is based on the determined cepstral coefficients."

Sharma et al. disclose performing segmentation based on the values of Kmax and Kmin that depend on cepstral coefficients. (elems. 20, 22, FIG. 1 and Col. 7, lines 23-28). Since, K max (maximum number of segments for each frame) is computed using SVF (Col. 6, line 29-31), which itself is a function of cepstral coefficients (Eq. 2),

the segmentation of each frame is ultimately performed based on the cepstral coefficients.

It would have been obvious to one of ordinary skill in the art at the time the invention was made to modify Juang et al. as taught by Sharma et al. in order to augment the segmentation process of frame data, because the detection of phoneme boundaries using cepstral coefficients would improve the speed and accuracy of the resulting segmentation process.

As per claim 18, Juang et al. discloses the method where readjusting scores are fed from block 360 back into block 325. (FIG.3) Thus the HMM and corresponding scores are updated using the weighting factor Wi (elem. 360, FIG. 3) which reflects the level of confidence in the current model. Therefore, the final recognition result (Elem. 355, FIG. 3) will incorporate the modification through weighting factor W.

As per claim 21, Juang et al. do not disclose "determining cepstral coefficients for the received frames of acoustic data, wherein the determining of the segmentation information is based on the determined cepstral coefficients."

Sharma et al. disclose performing segmentation based on the values of Kmax and Kmin that depend on cepstral coefficients. (elems. 20, 22, FIG. 1 and Col. 7, lines 23-28). Since, K max (maximum number of segments for each frame) is computed using SVF (Col. 6, line 29-31), which itself is a function of cepstral coefficients (Eq. 2),

the segmentation of each frame is ultimately performed based on the cepstral coefficients.

It would have been obvious to one of ordinary skill in the art at the time the invention was made to modify Juang et al. as taught by Sharma et al. in order to augment the segmentation process of frame data, because the detection of phoneme boundaries using cepstral coefficients would improve the speed and accuracy of the resulting segmentation process.

As per claim 26-27, Juang et al. do not disclose re-ordering the modified recognition hypothesis scores, and further using the re-ordered modified recognition hypothesis scores for the recognizing of the patterns in the frames of acoustic data.

However, it would have been obvious to one of ordinary skill in the art at the time the invention was made that modification of HMM scores (training) (as applied to claim 25) would necessarily involve re-ordering of scores of HMM outcomes.

6. Claims 31,32 are rejected under 35 U.S.C. 103(a) as being obvious over Muroi (4,918,731)

Muroi does not disclose processing unit for generating frame numbers and trainer/HMM decoder.

Muroi discloses:

- receiving frames of speech data (12, FIG. 1)

- the use of end frame numbers to compute duration of the speech pattern (phoneme segment) (Col. 7, lines 10-20). The duration of speech pattern is itself used in the calculations of weights for HMM state transitions (Col. 5, line 50). Therefore, frame numbers are used for calculation of weights for HMM state transitions.

It would have been obvious to one of ordinary skill in the art at the time the invention was made that the system of Muroi necessarily involved a processing unit that generated frame numbers and a trainer/HMM decoder for using the frame numbers to generate weights, since Muroi's system comprises hardware and software, and thus would require modules which would produce the corresponding frames numbers and also recognize patterns using weighted HMM transition probabilities.

Conclusion

7. The prior art made of record and not relied upon is considered pertinent to applicant's disclosure.

Fanty et al. (6,535,851) teaches speech segmentation based on cepstral coeffecients (6,535,851)

Ruey-Ching et al., "Improvement in Connected Mandarin Digit Recognition by Explicitly Modeling Coarticulatory Information" teaches that the number of peaks in cepstral

coefficients remains the same across the frames corresponding to the same segments. (p. 655)

Naylor et al (5,806,034) teach the system for backtracking in HMMs which relies on frame numbers. (FIG. 7)

Junqua (5,806,030) teaches clustering methods for HMM speech recognizers.

8. Any inquiry concerning this communication or earlier communications from the examiner should be directed to Dmitry Brant whose telephone number is (703) 305-8954. The examiner can normally be reached on Mon. - Fri. (8:30am - 5pm).

If attempts to reach the examiner by telephone are unsuccessful, the examiner's supervisor, Talivaldis Ivars Smits can be reached on (703) 306-3011. The fax phone number for the organization where this application or proceeding is assigned is (703) 872-9306.

Any inquiry of a general nature or relating to the status of this application or proceeding should be directed to Tech Center 2600 receptionist whose telephone number is (703) 305- 4700.

DB

4/7/04

DORIS H. TO
SUPERVISORY PATENT EXAMINER
TECHNOLOGY CENTER 2600